TAXONOMÍA DE INFORMACIÓN PERSONAL DE SALUD PARA GARANTIZAR LA PRIVACIDAD DE LOS INDIVIDUOS

Taxonomy of personal health information to ensure the privacy of individuals

Abel Lozoya-de-Diego, María-Teresa Villalba-de-Benito y María Arias-Pou

Abel Lozoya-de-Diego es ingeniero técnico en Informática de Gestión, ingeniero en Informática por la *Universidad de Valladolid* y máster universitario en Seguridad de las Tecnologías de la Información y las Comunicaciones por la *Universidad Europea de Madrid* en la que cursa el programa de doctorado. Ha complementado su formación con cursos tales como Mitigación de botnets del *Ccdcoe* de la *OTAN* y la certificación de Auditor Jefe en *ISO 27001*. Colabora como docente en el *Máster Universitario Oficial en Seguridad y Defensa* de la *Universidad Antonio de Nebrija*.
http://orcid.org/0000-0002-3225-8909

21041198@live.uem.es
abellozoya@gmail.com

María-Teresa Villalba-de-Benito es licenciada en Ciencias Matemáticas, especialidad de Ciencias de la Computación por la *Universidad Complutense de Madrid*, y doctora en Ingeniería Informática por la *Universidad de Alcalá* con premio extraordinario de doctorado. Profesora titular de Lenguajes y Sistemas informáticos en la *Universidad Europea de Madrid* desde 2002. Es directora del *Máster Univ. en Seguridad de Tecnologías de la Información y la Comunicación*. Acreditada por la *ACAP*. Representante española de *Legal & Security Issues Task Force* en *Cepis*.
http://orcid.org/0000-0003-0443-5979

maite.villalba@universidadeuropea.es

María Arias-Pou es licenciada en Derecho por la *Universidad de Navarra* y máster en Derecho de las Nuevas Tecnologías de la Información y las Comunicaciones, por *Icade*. Doctora en Derecho por la *Universidad Europea*. Directora de *Arias Pou Abogados TIC*. En septiembre de 2013 obtuvo la certificación *ACP-APEP* como consultor perfil jurídico mediante el procedimiento de *grandfathering*. Es asociada de *APEP* y coordinadora de su *Comisión de Menores*. Ha sido profesora del *Departamento de Derecho* en la *Universidad Europea*.
http://orcid.org/0000-0003-4613-4023

info@ariaspou.com

Universidad Europea de Madrid
Tajo, s/n. Villaviciosa de Odón (Madrid), España

Resumen

El crecimiento de los dispositivos de la internet de las cosas (*IoT*, *Internet of Things*) junto con su capacidad de almacenar grandes cantidades de datos de los usuarios, supone un nuevo reto a la privacidad del individuo. Un conocimiento de los datos médicos que son necesarios para los profesionales, supone un importante paso, pero deben ser protegidos y posteriormente anonimizados. Con el fin de avanzar hacia un modelo de datos de carácter médico protegidos específico del marco español, se ha realizado una revisión sistemática de la legislación sobre protección de datos a nivel internacional, así como de los trabajos relacionados. Posteriormente, con la ayuda de los actores involucrados, a través de metodologías cualitativas y cuantitativas se han obtenido los datos médicos más importantes clasificados según su relación, extrayendo además interesantes conclusiones sobre la importancia dada por los profesionales a los atributos clínicos utilizados.

Palabras clave

Protección de datos; Datos médicos; Privacidad; Anonimización; Seudonimización; *Big data*; Datos masivos; Internet de las cosas.

Artículo recibido el 10-09-2016
Aceptación definitiva: 21-11-2016

Abstract

The growth of the internet of things, along with its ability to store large amounts of user data, is a new challenge to individual privacy. Medical data is necessary for professionals, but these data have to be protected (PHI, Protected Health Information) and anonymized. In order to move towards a model of personal health information in Spain, a systematic review of the data protection legislation at an international level, as well as other related work, was carried out. Later, with the help of stakeholders, qualitative and quantitative methodologies were applied in order to obtain medical attributes classified according to their relationship. Interesting conclusions about the importance given by professionals to clinical attributes are described.

Keywords

Data protection; Medical data; Privacy; Anonymisation; Pseudonymisation; Big data; Internet of things; IoT.

Lozoya-de-Diego, Abel; **Villalba-de-Benito, María-Teresa; Arias-Pou, María** (2017). "Taxonomía de información personal de salud para garantizar la privacidad de los individuos". *El profesional de la información*, v. 26, n. 2, pp. 293-302.

https://doi.org/10.3145/epi.2017.mar.16

1. Introducción

En el último lustro numerosos estudios han demostrado un aumento progresivo en el número de ciberataques en el área de la salud con el fin de sustraer datos médicos individuales de los pacientes (**Villalba-de-Benito** *et al.*, 2015). El motivo fundamental es que estos datos médicos tienen un especial atractivo para los ciberdelicuentes dado que constan de una gran cantidad de información personal del individuo útil para ser utilizada con fines maliciosos. Además, diversos informes han puesto de manifiesto una falta de recursos y tecnologías para proteger dichos datos (*Ponemon Institute LLC*, 2015; *Symantec*, 2015). A esto hay que añadir también el aumento de las nuevas tecnologías de monitorización sobre la salud y/o el bienestar (también llamado internet de las cosas o *IoT* de las siglas en inglés) (**Gachet-Páez** *et al.*, 2015) como es el caso de la tecnología llevable o *wearable*. Estos dispositivos extraen diariamente una cantidad ingente de datos individuales que tendrían que ser sometidos a un nivel alto de protección según las leyes de protección de datos. Posteriormente estos datos se almacenan en entornos *cloud* (en la nube) y se analizan usando tecnologías de datos masivos o *big data* con el fin de extraer conocimiento. Se estima que la cantidad de datos personales generados por individuos aumente de manera progresiva y a un ritmo considerable, debido al gran interés tanto de la industria, como de los gobiernos en esta tecnología por su capacidad para extraer conocimiento (**Chen**; **Mao**; **Liu**, 2014). Además, la cantidad de datos en abierto o compartidos por la comunidad científica, así como, movimientos como *open access* u *open data*, favorece dicho crecimiento (**Nina-Alcocer**; **Blasco-Gil**; **Peset**, 2013).

> En el caso de España, la *LOPD* no contempla los datos médicos que deben ser anonimizados aun siendo datos clasificados de nivel alto de protección

Por otra parte, tanto los dispositivos *IoT*, como los proveedores *cloud* disponen de políticas de privacidad en las que no siempre se especifica qué datos se envían y almacenan en la nube, o qué se hace con los datos del usuario. Otras veces, cuando se especifica, se cometen infracciones graves contra el derecho al acceso y rectificación de los datos por parte del usuario, no permitiendo, por ejemplo, la objeción del mismo si los datos son enviados a un tercero (**Villalba-de-Benito** *et al.*, 2015). Por tanto, los usuarios de estos dispositivos se encuentran indefensos, no teniendo conocimiento de la gran cantidad de información personal que están compartiendo, ni de la forma de modificar o eliminar la información personal que no quieran compartir.

En el caso de la salud, existe información que debe ser protegida para mantener el derecho de privacidad del paciente. En Estados Unidos, dicha información ha sido denominada *Protected Health Information* (PHI) (*United States*, 1996). Tal como declaró la *American Medical Informatics Association* (*AMIA*) (**Hripcsak** *et al.*, 2014), la protección de los pacientes como grupo de individuos vulnerables debe ser hecha mediante técnicas de anonimización y de privacidad desde el diseño (**Fischer**; **Morte-Ferrer**, 2013). La anonimización consiste en desvincular la información personal del individuo, de manera que dificulte, en la medida de lo posible, el proceso de reidentificación. La legislación española relativa a protección de datos, a diferencia de la de otros países, no contempla los datos médicos que deben ser anonimizados por ser considerados de nivel alto (clasificación según *Ley orgánica 15/1999, de 13 de diciembre, de Protección de datos de carácter personal* (*España*, 1999). Esto dificulta en gran medida un proceso ya de por sí complejo. Una opción que se podría plantear en relación a qué campos anonimizar, podría ser la realización de una anonimización del conjunto completo de los atributos clínicos, sin embargo, esto podría derivar en una pérdida de información relevante a la hora de extraer conocimiento de los datos recogidos de los pacientes. Por este motivo es necesario realizar una correcta elección de los datos que deben ser anonimizados, de manera que se pueda extraer conocimiento de los mismos garantizando el derecho a la protección de datos del paciente.

El *Grupo de trabajo del Artículo 29* de la Directiva 95/46/CE (*Europa*, 1995), publicó el 10 de abril de 2014 su *Opinión 05/2014* sobre técnicas de anonimización (*Europa*, 2014). En dicha *Opinión* se reconoce la inexistencia de una técnica

segura, aunque sí que se especifica que una estrategia que evite los riesgos de singularización, "vinculabilidad"[1] e inferencia, mantendrá la solidez necesaria para impedir la reidentificación de los datos mediante los medios más probables y razonables. Así mismo, este grupo de trabajo indica que: "han de definirse con claridad los requisitos previos (el contexto) y los objetivos del proceso para obtener la anonimización deseada al mismo tiempo que se generan datos útiles". Además recomienda que "la solución óptima debe decidirse caso por caso y puede conllevar la combinación de diversas técnicas". Este es un proceso largo y, en consecuencia, costoso. Por este motivo, si en la decisión de cada proceso de anonimización la persona responsable dispusiera del conjunto de atributos para ese dominio de conocimiento a ser anonimizados, podría reducirse significativamente el tiempo dedicado al proceso de anonimización y, en consecuencia, los costes, además de limitar la pérdida de información debido a la anonimización de datos que no es necesario anonimizar.

https://www.agpd.es/portalwebAGPD/internacional/Europa/grupo_29_europeo/index-ides-idphp.php

En este trabajo se desarrolla una caracterización de los datos clínicos a anonimizar con el fin de mantener la privacidad del paciente, incluido el caso de investigaciones que hagan uso de técnicas novedosas como *big data*. Un tema vulnerable frente a los accesos o consultas del personal sanitario en general, y en la utilización de técnicas de *big data* en particular.

Este artículo se organiza de la siguiente manera:

La sección 2 muestra un resumen del marco regulatorio; la sección 3 presenta la metodología utilizada en este trabajo y el análisis de los datos obtenidos; y la sección 4 presenta las conclusiones del estudio.

2. Marco regulatorio

En esta sección se recogen las referencias normativas tanto a nivel nacional, comunitario e internacional sobre protección de datos distinguiendo entre ellas de un lado, las que protegen los datos de salud y las que no, y, de otro lado, las que definen los atributos médicos que deben ser objeto de anonimización de los tratamientos masivos de datos con el objetivo de favorecer la investigación y evitar los tratamientos innecesarios de datos. Se pone así de relieve cómo aquellas normativas que detallan los atributos médicos a anonimizar aportan un mayor grado de seguridad jurídica al desarrollo del tratamiento masivo de datos y a la internet de las cosas.

En Europa, el recientemente aprobado *Reglamento general de protección de datos*, de 26 de abril de 2016, (en adelante *RGPD*) (*Unión Europea*, 2016) define en su artículo 4 los datos relativos a la salud como los datos personales relativos a la salud física o mental de una persona, incluida la prestación de servicios de atención sanitaria, que revelen información sobre su estado de salud. Además, el *RGPD* refleja en ese mismo artículo el concepto de seudonimización como técnica para el tratamiento de datos personales que garanticen que dichos datos no puedan atribuirse a un interesado en particular sin recurrir a información adicional. Los datos que puedan atribuirse a una persona física mediante la utilización de información adicional deben considerarse como información sobre una persona física identificable y su tratamiento queda sometido a la protección de datos. Para determinar si una persona es identificable el *RGPD* dispone que deben tenerse en cuenta todos los medios a su alcance, por ejemplo, la singularización, que pudiera utilizar el responsable del tratamiento o cualquier otro individuo para identificar directa o indirectamente a dicha persona. Para determinar si existe una probabilidad razonable en la utilización de unos medios determinados para la identificación de una persona física deben tenerse en cuenta todos los factores objetivos, como los costes y el tiempo necesarios para la identificación. Para ello se debe tomar en consideración tanto la tecnología disponible en el momento del tratamiento como los avances tecnológicos. Y a continuación dispone que, por lo tanto, los principios de protección de datos no deben aplicarse a la información anónima (información que no guarda relación con una persona física identificada o identificable), o a los datos convertidos en anónimos de forma que el interesado a quien se refieren no sea, o ya no resulte, identificable. En consecuencia, el Reglamento no afecta al tratamiento de dicha información anónima, ni a los relacionados con fines estadísticos y de investigación pero sí a los datos seudonimizados que son los que regula.

Otras referencias que encontramos en el *RGPD* a seudonimización son el artículo 32 que se refiere a la seguridad de los datos. Éste, dispone que teniendo en cuenta el estado de la técnica, los costes de aplicación, la naturaleza, el alcance, el contexto, los fines del tratamiento, así como el resto de probabilidad y gravedad variables, para los derechos y libertades de las personas físicas, el responsable y el encargado del tratamiento aplicarán medidas técnicas y organizativas apropiadas para garantizar un nivel de seguridad adecuado al riesgo. Ello incluiría incluso y según corresponda, la seudonimización y el cifrado de datos personales.

Por su parte, el artículo 40 del *RGPD* dedicado a los códigos de conducta recoge una referencia expresa a la seudonimización de los datos. Así, indica que las asociaciones y otros organismos representativos de categorías de responsables o encargados del tratamiento, podrán elaborar códigos de conducta o modificar/ampliar dichos códigos con el propósito de especificar la aplicación de disposiciones del Reglamento, por ejemplo, la seudonimización de datos personales.

> "Los modelos cualitativos en el ámbito de protección de datos médicos pueden ser instrumentos de gran utilidad en la protección de datos relacionados con la salud"

En materia de protección de datos a nivel nacional de España se debe atender a la *Ley orgánica 15/1999*, de 13 de diciembre, de *Protección de Datos de Carácter Personal*, en adelante *LOPD* (*España*, 1999), y al *Real decreto 1720/2007*, de 21 de diciembre, *por el que se desarrolla el Reglamento de la LOPD*, en adelante *RDLOPD* (*España*, 2007). La *LOPD* no proporciona ningún concepto que comprenda el campo de datos de la salud a un nivel tal que permita identificarlos para seguir un proceso de anonimización. Por su parte, el *RDLOPD* en su artículo 5.1.g define los datos de carácter personal relacionados con la salud como "las informaciones concernientes a la salud pasada, presente y futura, física o mental, de un individuo. En particular, se consideran datos relacionados con la salud de las personas los referidos a su porcentaje de discapacidad y a su información genética". Aunque la definición de datos de la salud no acaba de indicar qué tipo de campos de información son considerados datos relacionados con la salud, cita dos posibles categorías:

a) los atributos relacionados con la discapacidad

b) la información genética.

Debido a la falta de identificación de atributos médicos que deberían ser anonimizados en la normativa aplicable a nivel europeo y nacional, hemos llevado a cabo un análisis de la normativa internacional en protección de datos para conocer los datos médicos que son protegidos en otros países con el fin de obtener el conjunto total de datos médicos potencialmente anonimizables y analizarlos para adaptarlos al caso español. Tras el análisis de la legislación en esta materia a nivel internacional se obtuvieron 122 atributos médicos.

En África las legislaciones de protección de datos que han aportado nuevos atributos médicos al estudio han sido Sudáfrica con su ley *Electronic communications and transactions act*, *2002* aprobada el 31 de julio de 2002 (*South Africa*, 2002) y Marruecos con *Loi nº 09-08 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel*, de 18 de febrero de 2009 (*Maroc*, 2009b) y su desarrollo con el *Décret nº 2-09-165 du 25 joumada I 1430 (21 mai 2009) pris pour l'application de la loi nº 09-08 relative à la protection des personnes physiques à l'égard des traitements des données à caractère personnel*, de 21 de mayo de 2009 (*Maroc*, 2009a).

En Europa, Finlandia, Lituania e Islandia han aportado campos médicos protegidos que no habían sido contemplados en la *Directiva 95/46/CE* (*Europa*, 1995), que ha sido derogada por el *RGPD* con efecto a partir del 25 de mayo de 2018. En Finlandia la ley *Personal data act 523/1999* (*Finland*, 1999) en su capítulo 3 sección 11 indica qué tipo de datos médicos se deben proteger: el estado de la salud, enfermedad o discapacidad y tratamientos. En Lituania la ley *Law on legal protection of personal data n. I-1374* (*Lithuania*, 1996) especifica los datos de la salud descomponiéndolos en el estado, diagnóstico, pronóstico y tratamiento. Islandia, aunque no pertenece a la Unión Europea, en su normativa *Act on the protection of privacy as regards the processing of personal data* (*Iceland,* 2000) incluye datos de consumo de alcohol, de narcóticos y de drogas.

En Oceanía, Australia dispone de la ley *Privacy act n. 119 of 1988* (*Australia*, 1988) en la que se han ido realizando modificaciones hasta el 12 de marzo del 2014. En esta ley quedan definidos los campos que pertenecen a la información identificable sobre un individuo y un conjunto de datos que se considera como información médica.

En América, se evidencia una gran diferencia entre la normativa relativa a protección de datos de Norteamérica y la del resto del continente. Las leyes de protección de datos de Canadá:

a) *Privacy act* de 1985 (*Canada*, 1985) y

b) *Personal information protection and electronic documents act*, de 1 de junio de 2000 (*Canada*, 2000);

En los Estados Unidos, con la *Health insurance portability and accountability act* (conocida como *Hipaa*), de 21 de agosto de 1996 (*United States*, 1996), se definieron qué datos son los que están englobados en la categoría de datos personales. En ese país se proporciona un listado de campos de información médica que deben ser protegidos. En América Central y América del Sur, la definición tiene un nivel de abstracción significativo tal y como ha sucedido en Europa con la *Directiva 95/46/CE* (*Europa*, 1995). Sin embargo, el país Trinidad y Tobago con su A*ct No. 13* of 2011, aprobada el 3 de junio de 2011 (*Trinidad and Tobago*, 2011) y el estado mexicano de Nuevo León con la *Ley de transparencia y acceso a la información del Estado de Nuevo León* con una última reforma de 17 de septiembre de 2012 (*Nuevo León*, 2008), incluyen unos atributos médicos dentro de la definición de datos personales de la salud.

Tras realizar este análisis exhaustivo de la normativa nacional e internacional sobre protección de datos, se han obtenido las siguientes conclusiones:

- Existen países con leyes de protección de datos personales que no contemplan los datos de salud como datos a proteger. Sin embargo, son una minoría a nivel mundial.
- Hay una gran mayoría de países que en sus leyes de protección de datos han reutilizado el concepto de protección de datos de la salud de otras instituciones. Sin embargo, no han definido qué datos personales se consideran como datos personales sanitarios. Entre ellos destacamos España con su *LOPD*.
- Existe un número reducido de países que incluyen en su ley de protección de datos aquellos datos personales relativos al campo de la salud. Además, incluyen una serie de campos que deben ser protegidos dentro del campo sanitario para garantizar la protección del individuo frente a su identificación.

3. Metodología de investigación

Una taxonomía es una clasificación construida sobre una base sólida de revisión del marco teórico y empíricamente validada (**Lai**; **Cheng**; **Yeung**, 2004). Para validar empíricamente las taxonomías suele utilizarse el análisis de datos multifactorial, ya que permite reconocer las similitudes entre cada una de las variables o atributos, y agruparlos en categorías. Por ello, en este trabajo se ha realizado una revisión sistemática de los estudios relacionados que, posteriormente se ha validado con expertos utilizando técnicas cualitativas y cuantitativas.

En primer lugar, el estudio de la legislación y de los trabajos dio lugar a un conjunto inicial de 122 atributos médicos (**Lozoya-de-Diego**; **Villalba-de-Benito**; **Arias-Pou**, 2016). Dado que estos atributos se recogieron de trabajos internacionales, se llevó a cabo un proceso de adaptación y filtrado de los mismos a través: primero de una revisión interna para identificar duplicados y renombrar cuando fuera necesario y, posteriormente de posibles aspectos clínicos mediante cualitativos y cuantitativos. Con el fin de obtener una visión más amplia y resultados que reflejaran de forma más precisa la visión de todos los actores involucrados en el proceso, se recogió la perspectiva tanto de personal sanitario, como de personal informático trabajando en el área de la salud.

Para apoyar el proceso de revisión interna de atributos se utilizó el *Real decreto 1093/2010*, de 3 de septiembre (*España*, 2010), por el que se aprueba el conjunto mínimo de datos de los informes clínicos en el Sistema Nacional de Salud. Este *Real decreto* en su artículo 3 y en los anexos aporta el listado de campos médicos que como mínimo deben contener una serie de documentos clínicos, lo que permitió apoyar la decisión sobre qué atributos son utilizados a nivel nacional.

En la revisión interna se llevó a cabo un proceso de consultas a expertos tanto cualitativas como cuantitativas. En primer lugar, se utilizaron técnicas cualitativas realizando entrevistas estructuradas a personal sanitario, personal informático involucrado en procesos médicos e investigadores del área. Los entrevistados eran:

- cuatro médicos (uno de medicina interna, un forense, un psicólogo y un urólogo);
- dos investigadores del área de informática aplicada a la salud, y
- un informático del área de salud.

Tras esta revisión interna, el número de atributos clínicos se redujo a 88.

Finalmente, con el fin de confirmar los atributos y obtener los esenciales, se llevó a cabo una investigación cuantitativa usando como instrumento el cuestionario para la recogida de información sobre los atributos médicos. Dado que el objetivo era confirmar los atributos se llevaron a cabo diferentes análisis estadísticos basados en el estudio de las correlaciones entre las variables. Además, se comprobó la fiabilidad del cuestionario para validar la consistencia interna del mismo. Posteriormente se aplicó análisis multifactorial con el fin de reducir las redundancias de los atributos bajo análisis y obtener una clasificación de los mismos. En las siguientes secciones se presentan las fases del proceso de investigación utilizados en este trabajo.

> El modelo propuesto puede ser incorporado en tecnologías de análisis masivas de datos, tales como *big data*, para facilitar de forma significativa la manipulación de la ingente cantidad de datos

3.1. Recogida de datos

Tomando como base el conjunto de 88 atributos médicos ya depurado, se elaboró un cuestionario con el propósito de recoger la opinión de los expertos para su posterior análisis estadístico. El cuestionario constaba de dos partes: una sección inicial en la que se explicaba el objetivo del estudio y se recopilaban los datos demográficos de los encuestados con el fin de confirmar el nivel de conocimiento de los profesionales invitados a participar en el estudio; la sección principal con las preguntas asociadas a los atributos médicos obtenidos de la revisión bibliográfica tras el análisis interno y el estudio preliminar con expertos. Dado el gran número de atributos, con el fin de hacer más comprensible al encuestado esta sección, los atributos se clasificaron en categorías. Dichas categorías fueron validadas por el grupo de expertos a través del análisis cualitativo resultando finalmente en las siguientes: datos personales, datos de dispositivos implantados y datos médicos. Se utilizó una escala de 5 puntos tipo Likert desde 1 (nunca o casi nunca) a 5 (siempre). Al final de todas las preguntas se disponía de un espacio en blanco para que el participante pudiera sugerir atributos médicos no recogidos en el cuestionario y que considerase de importancia en el estudio.

Debido a que una parte importante del proceso está basada en la recogida de los datos, se deben controlar los perfiles de los profesionales invitados a complementar el cuestionario. Por este motivo los cuestionarios se elaboraron en dos aplicaciones web con acceso restringido y con garantías de anonimato para los participantes. Los encuestados fueron invitados específicamente a instancias de su perfil durante eventos, desde los contactos obtenidos del *Catálogo Nacional de Hospitales* del *Ministerio de Sanidad, Servicios Sociales e Igualdad* (*España*, 2015) y con la colaboración de las asociaciones profesionales como la *Asociación de Técnicos*

de Informática o la *Asociación Profesional Española de Privacidad*. Para excluir los cuestionarios de los participantes que no cumplieran con el perfil sanitario, se incluyeron los datos demográficos.

El proceso de recogida de datos recopiló un total de 169 cuestionarios de los cuales 120 eran cuestionarios completos y válidos por lo que el margen de error fue del 7,5% con un intervalo de confianza del 90%. La mayoría de los encuestados tenían entre 36 y 49 años (42,37%) o de 50 a 65 años (33,89%). El sector más representado por los encuestados el de salud pública (55,93%) seguido del sector de educación y formación (12,71%). La muestra estaba homogéneamente balanceada en relación a la especialidad médica de los profesionales de la salud. Además, los profesionales contaban con alta experiencia en su trabajo (41,52%). En relación a la protección de datos, la mayoría de los encuestados tenía un nivel de familiaridad de usuario/paciente (64,41%); disponiendo el 42,37% de formación a nivel básico en esta materia y el 27,12% nivel alto o de responsable de ficheros de datos. Por último, casi la mitad de los profesionales encuestados se dedicaba al cuidado de pacientes (49,15%), seguidos por personal dedicado a la investigación en el área de la salud (14,41%).

Además, se incluyó una pregunta adicional al cuestionario con el fin de saber, desde el punto de vista de los expertos, la importancia de disponer de un modelo de atributos con datos médicos personales que puedan vulnerar la privacidad de los pacientes/usuarios. Los resultados mostraron que para el 51% de los encuestados es muy importante disponer de un modelo de este tipo, mientras que el 32% consideran que es indispensable. Podemos, por tanto, concluir en la alta utilidad de disponer de un modelo de atributos médicos personales que contengan información sensible del usuario/paciente a proteger a partir del cual los profesionales de la salud puedan determinar cuáles son realmente necesarios y deben ser protegidos, y cuáles no y, por tanto, deben eliminarse.

> No se debe ignorar el uso y aplicación de técnicas de ciberdefensa en tecnologías de tratamiento masivo de datos para minimizar los riesgos asociados a la fuga de información

3.2. Análisis de los datos

El análisis estadístico se llevó a cabo con el programa SPSS versión 22. En primer lugar como medida de control se hizo un análisis exploratorio con el fin de detectar posibles errores durante la recogida de datos, así como para comprobar la viabilidad del análisis factorial. Posteriormente se hizo un análisis descriptivo de todas las variables o atributos médicos del estudio (medias, desviación estándar, mediana, mínimo y máximo y las frecuencias absolutas y relativas). Además, se realizó una revisión de los diagramas de caja para determinar los errores de entrada de datos y el coeficiente de variación para comprobar la homogeneidad de los datos. Como resultado de este análisis se eliminaron 37 atributos médicos por tener todos ellos una baja importancia según los expertos (media inferior a 2, mediana menor o igual a 2 y moda de 1).

Posteriormente se llevó a cabo una exploración preliminar de las relaciones entre los atributos a través del análisis de la matriz de correlación usando el coeficiente de Pearson. Se considera que existe correlación entre las variables, cuando los valores del coeficiente de Pearson son superiores a 0,30 y los determinantes de matrices cercanos a 0 (**Hair** *et al*., 1998). En nuestro estudio, el análisis de la matriz de correlación resultó en variables correladas al arrojar valores superiores a 0,30 del coeficiente de Pearson y determinantes con valores positivos y cercanos a 0. Por lo tanto, se confirma que se puede aplicar el análisis factorial.

> Se ha obtenido una taxonomía de los datos personales a proteger sobre la salud y bienestar de los individuos, basada en los marcos legales internacionales y artículos y trabajos relacionados

Con el objetivo de verificar la fiabilidad del cuestionario se llevó a cabo un análisis del coeficiente alfa de Cronbach, calculándolo tanto para el total del cuestionario, como para cada una de las categorías o dimensiones consideradas. Dicho análisis permite verificar que todos los factores (atributos médicos, en nuestro caso) miden el mismo concepto, así como si los resultados son precisos y consistentes. Convencionalmente el mínimo utilizado para el coeficiente alpha de Cronbach como aceptable está establecido como 0,7, siendo considerados los valores por encima de 0,8 como buena fiabilidad, y 0,9 como excelente (**Tabachnick**; **Fidell**, 2006). Como se puede comprobar en la tabla 1, para todas las variables los coeficientes alfa están comprendidos entre 0,880 y 0,964, demostrando así, una alta consistencia y fiabilidad.

Por otra parte, la validez del constructo es el grado en el que el instrumento (cuestionario) mide aquello para lo que fue diseñado. Para medir la validez se utilizó análisis factorial exploratorio que, además, permite identificar la estructura subyacente de las relaciones entre las variables, obteniendo así una validación predictiva del modelo. Antes de aplicar dicho análisis es necesario verificar que se cumplen las condiciones para ello. El índice Kaiser-Meyer-Oklin, o KMO, y el test de esfericidad de Bartlett son las medidas utilizadas en este caso. El índice KMO mide la relación entre las variables. El valor mínimo comúnmente aceptado para dicho índice es 0,5, considerando buenos valores a aquellos entre 0,7 y 0,8, y meritorios los superiores a 0,8 (**Punter**; **Solingen**; **Trienekens**, 1997). En relación con el test de esfericidad de Bartlett el modelo es significativo para p-valores menores de 0,05. Como puede verse en la tabla 1, los valores de KMO se encuentran entre 0,807 y 0,826 y el test de esfericidad de Bartlett es en todos los casos inferior al valor mínimo de referencia. En ese caso, se concluye por tanto que puede aplicarse el análisis multifactorial.

La tabla 1 muestra el resultado final tras aplicar el análisis factorial exploratorio. Como el objetivo era reducir los atri-

Tabla 1. Resumen de análisis factorial exploratorio

Categoría	Factor	Alpha Cronbach	KMO	Bartlett's test
Datos personales	**Datos de localización**	0,880	0,807	Sig 0,000 gl 105 chi.square 1.962,729
	Ciudad			
	Municipio			
	Comunidad autónoma			
	Datos de envejecimiento			
	Género			
	Fecha de nacimiento			
	Nacionalidad			
	Edad o elemento indicativo de edad			
	Sexo			
	Datos de contacto			
	Nombre			
	Primer apellido			
	Segundo apellido			
	Teléfono			
	Datos de procedencia			
	País de nacimiento			
	Origen étnico			
	Profesión			
Datos médicos	**Atención hospitalaria**	0,964	0,826	Sig 0,000 gl 630 chi-square 5.035,253
	Tratamiento			
	Alergias			
	Diagnosis			
	Enfermedades			
	Estado de embarazo			
	Estado de enfermedad			
	Historial clínico			
	Información recolectada durante el servicio sanitario			
	Intervenciones quirúrgicas			
	Número de historia clínica			
	Fecha de alta			
	Fecha de tratamiento			
	Fecha de visita			
	Estado mental			
	Centro de salud			
	Consumo de sustancias tóxicas			
	Información psicológica			
	Etiología y patogénesis			
	Incapacidades médicas			
	Estado de discapacidad			
	Salud mental (otra información mental no recogida en anteriores campos)			
	Códigos			
	Prótesis			
	Códigos de diagnosis			
	Códigos de tratamiento			
	Fecha de admisión			
	Ensayos clínicos			
	Vacunas			
	Número de identificación			
	Número de registros de ensayos clínicos			
	Analíticas			
	Servicio prestado			
	Laboratorios que analizaron las pruebas médicas			
	Fecha de recogida de muestras			
	Otra información médica			
	Medidas corporales (talla, peso)			
	Prognosis			
	Residencias de ancianos			
	Fecha de deceso			
	Salud física (otra información física no recogida en anteriores campos)			
	Términos médicos que identifican patologías no habituales			
TOTAL		0,964	0,712	Sig 0,000 gl 1275 Chi-square 8.329,498

Tabla 2. Taxonomía de datos médicos de carácter personal

Dominios	Dimensiones
Datos personales	Datos de ubicación
	Datos de envejecimiento
	Datos de contacto
	Datos de procedencia
Datos médicos	Atención hospitalaria
	Estado mental
	Códigos
	Ensayos clínicos
	Analíticas
	Otra información médica

butos eliminando duplicidades, se aplicó rotación Varimax con normalización de Kaiser y el criterio de los eigenvalues (**Hair** *et al*., 1998) para decidir qué factores o atributos mantener. El valor de corte utilizado para mantener un factor fue 0,32 (**Tabachnick**; **Fidell**, 2006). Como resultado de ello, las categorías iniciales fueron subdivididas a su vez en otras categorías o componentes que a su vez fueron renombradas según los atributos que contenían, dando lugar a la taxonomía de atributos médicos personales para el caso de España que se muestra en la tabla 2. Así, la taxonomía queda definida por 2 dominios, datos personales y datos médicos, y cada uno de ellos, por 4 y 6 dimensiones.

> Se requiere de una concienciación de los profesionales de la salud y usuarios finales sobre la importancia de proteger la información personal de salud

4. Conclusión

Los modelos cualitativos en el ámbito de la protección de datos médicos pueden ser instrumentos de gran importancia en la protección de datos. Sin embargo, la caracterización, diseño, adaptación y desarrollo de un modelo cualitativo que simule el comportamiento de un proceso de anonimización de atributos clínicos no es un trabajo sencillo dada la gran cantidad de factores involucrados en el mismo, como son las legislaciones existentes, zonas geográficas, tipos de pacientes, tecnologías utilizadas o umbrales biométricos. Ello hace fundamental disponer de una taxonomía con los datos médicos a proteger adaptados al país o zona geográfica en donde va a ser aplicado.

En base al amplio conjunto de referencias de la bibliografía se ha conceptualizado, construido, refinado y validado una taxonomía de atributos clínicos. Para ello, se ha recogido la opinión de expertos acerca de los atributos utilizados a nivel nacional y, a continuación, se han analizado estos datos. Como resultado hemos obtenido una taxonomía que permite reducir un conjunto inicial de atributos médicos, así como identifica aquellos que tienen que someterse a procesos de anonimización para mantener la privacidad de los pacientes en el ámbito de la salud. Este diseño ha permitido pasar de un escenario de 9 dimensiones y 122 items, a 2 dimensiones y 88 items. Por consiguiente, se ha logrado una reducción de un 27,86% de las variables. Con ello, se ha obtenido una taxonomía de atributos médicos que deben ser protegidos mediante técnicas de anonimización. La utilización de dicha taxonomía permitirá no sólo mejorar la privacidad del paciente sino, además, mejorar el tiempo de análisis de cada proceso de anonimización propuesto por el *Grupo de trabajo del Artículo 29*, así como limitar la pérdida de información relevante debida a una mala elección de los atributos a anonimizar.

Otra posible aplicación de la taxonomía aquí presentada es la de la concienciación del personal de salud y del usuario en general sobre la importancia de mantener la privacidad de los datos personales. La extensión de tecnologías *big data* e *IoT* hacen que el entorno en el que estos datos son manipulados sea cada vez menos restringido, lo que conlleva a un peligro de privacidad para el usuario de datos catalogados como de nivel alto de protección. Se hace necesario, por tanto, una mayor transparencia por parte de los proveedores de estos servicios y un empoderamiento digital del usuario, dotándole de herramientas que le permitan conocer el grado de información que está compartiendo en cada momento, así como la posibilidad de modificación y eliminación de esos datos. La Unión Europea, por su parte, ya está tomando medidas al respecto a través del *Grupo de trabajo del Artículo 29* (*Europa*, 2014) enviando notificaciones formales a los proveedores para que se ajusten a las leyes de protección de datos, como es el caso de *Microsoft Windows 10* en Francia (*France*, 2016).

Por último, resaltar el valor de la taxonomía propuesta de atributos clínicos como pilar base para la creación de futuros modelos más robustos y amplios no sólo para la protección de datos clínicos, sino para cualquier otra área relacionada con las tecnologías de uso masivo de datos.

Nota

1. Vinculabilidad. Palabra extraída del documento aunque no existe en español (*Grupo de trabajo sobre protección de datos del artículo 29*, 2014).

5. Bibliografía

Australia (1988). "Privacy act n. 119 of 1988". *Gazette 1988 n. S399 of 12 March 2014*.
https://www.legislation.gov.au/Details/C2016C00838

Canada (2000). "Personal information protection and electronic documents act". *Minister of Justice of 13 of April of 2000*. The consolidation is current to March 28, 2016.
http://laws-lois.justice.gc.ca/PDF/P-8.6.pdf

Canada (1985). "Privacy act". *Minister of Justice* current to April 12, 2016.
http://laws-lois.justice.gc.ca/PDF/P-21.pdf

Chen, Min; **Mao, Shiwen**; **Liu, Yunhao** (2014). "Big data: A survey". *Mobile networks and applications*, v. 19, n. 2, pp. 171-209, ISSN: 1572-8153
https://dx.doi.org/10.1007/s11036-013-0489-0

España (1999). "Ley orgánica 15/1999, de 13 de diciembre,

de Protección de datos de carácter personal". *Boletín oficial del estado* [BOE-A-2003-23936].
https://www.boe.es/buscar/act.php?id=BOE-A-1999-23750

España (2007). "Real decreto 1720/2007, de 21 de diciembre, por el que se aprueba el Reglamento de desarrollo de la Ley orgánica 15/1999, de 13 de diciembre, de protección de datos de carácter personal". *Boletín oficial del estado* [BOE-A-2008-979].
https://www.boe.es/buscar/act.php?id=BOE-A-2008-979

España (2010). "Real decreto 1093/2010, de 3 de septiembre, por el que se aprueba el conjunto mínimo de datos de los informes clínicos en el Sistema Nacional de Salud". *Boletín oficial del estado* [BOE-A-2010-14199].
http://www.boe.es/diario_boe/txt.php?id=BOE-A-2010-14199

España (2015). "Catálogo nacional de hospitales". *Ministerio de Sanidad, Servicios Sociales e Igualdad, Gobierno de España.*
http://www.msssi.gob.es/ciudadanos/prestaciones/centrosServiciosSNS/hospitales/home.htm

Europa (1995). "Directiva 95/46/CE del Parlamento Europeo y del Consejo de 24 de octubre de 1995 relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales ya la libre circulación de estos datos". *Diario oficial de las comunidades europeas L281 nº 38 de 23 de noviembre de 1995*. ISSN: 1012-9200.
https://goo.gl/iaq3dd

Europa (2014). "Opinión 05/2014 sobre técnicas de anonimización". *Grupo de trabajo sobre protección de datos del artículo 29.*
http://ec.europa.eu/justice/data-protection/article-29/documentation/opinion-recommendation/files/2014/wp216_es.pdf

Finland (1999). "Personal data act (523/1999)". *Finlex data bank.*
http://www.finlex.fi/en/laki/kaannokset/1999/en19990523.pdf

Fischer, Philipp E.; **Morte-Ferrer, Ricardo** (2013). "Big data: a challenge for data protection". *Huygens Editorial*, pp. 205-222, ISBN: 978 84 695 8160 5

France (2016). "Windows 10: CNIL publicly serves formal notice to Microsoft Corporation to comply with the French data protection act within three months". *Chair of the National Data Protection Commission (CNIL).*
https://www.cnil.fr/en/windows-10-cnil-publicly-serves-formal-notice-microsoft-corporation-comply-french-data-protection

Gachet-Páez, Diego; **De-Buenaga-Rodríguez, Manuel**; **Puertas-Sanz, Enrique**; **Villalba, María-Teresa**; **Muñoz-Gil, Rafael** (2015). "Big data processing using wearable devices for wellbeing and healthy activities promotion". *Springer International Publishing*. Ambient assisted living. ICT-based solutions in real life situations: *7th International Work-Conference*, Iwaal 2015, Puerto Varas, Chile, December 1-4, v. 9455, pp. 196-205.
https://doi.org/10.1007/978-3-319-26410-3_19

Hair, Joseph F.; **Anderson, Rolph E.**; **Tatham, Ronald L.**; **Black, William C.** (1998). "Multivariate data analysis" (5th ed.). Prentice Hall. ISBN: 978 0138948580.

Hripcsak, George; **Bloomrosen, Meryl**; **FatelyBrennan, Patti**; **Chute, Christopher G.**; **Cimino, Jim**; **Detmer, Don E.**; **Edmunds, Margo**; **Embi, Peter J.**; **Goldstein, Melissa M.**; **Hammond, William-Ed**; **Keenan, Gail M.**; **Labkoff, Steve**; **Murphy, Shawn**; **Safran, Charlie**; **Speedie, Stuart**; **Strasberg, Howard**; **Temple, Freda**; **Wilcox, Adam B.** (2014). "Health data use, stewardship, and governance: ongoing gaps and challenges: a report from AMIA's 2012 Health policy meeting". *Journal of the American Medical Informatics Association,* v. 21, n. 2, pp. 204-211.
http://jamia.oxfordjournals.org/content/21/2/204
http://dx.doi.org/10.1136/amiajnl-2013-002117

Iceland (2000). *Act on the protection of privacy as regards the processing of personal data*, n. 77/2000, May 10.
http://www.personuvernd.is/information-in-english/greinar/nr/438

Lai, Kee-Hung; **Cheng, T. C. Edwin**; **Yeung, Andy C. L.** (2004). "An empirical taxonomy for logistics service providers". *Maritime economics & logistics*, v. 6, n. 3, pp. 199-219.
https://dx.doi.org/10.1057/palgrave.mel.9100109

Lithuania (1996). "Law on legal protection of personal data of June 11". *World intellectual property organization,* n. I-1374 (new version of February 1, 2008, Law n. X-1444).
http://www.wipo.int/wipolex/en/text.jsp?file_id=202094

Lozoya-de-Diego, Abel; **Villalba-de-Benito, María-Teresa**; **Arias-Pou, María** (2016). "Análisis sobre la heterogeneidad en la legislación de protección de datos personales de carácter médico". *Diario La Ley* n. 8688, ISSN: 1989-6913

Maroc (2009). "Décret nº 2-09-165 du 25 joumada I 1430 (21 mai 2009) pris pour l'application de la loi nº 09-08 relative à la protection des personnes physiques à l'égard des traitements des données à caractère personnel". *Bulletin officiel* n. 5744 de 18-6-2009.
http://www.cdvm.gov.ma/sites/default/files/Decret_n209165.pdf

Maroc (2009). "Loi nº 09-08 relative à la protection des personnes physiques à l'égard du traitement des données à caractère personnel". *Bulletin officiel*, n. 5714 de 5-3-2009.
https://goo.gl/AQ86Pn

Nina-Alcocer, Víctor; **Blasco-Gil, Yolanda**; **Peset, Fernanda** (2013). "Datasharing: guía práctica para compartir datos de investigación". *El profesional de la información*, v. 22, n. 6, pp. 562-568.
https://doi.org/10.3145/epi.2013.nov.09

Nuevo León (2008). "Ley de transparencia y acceso a la información del Estado de Nuevo León". *Periódico oficial* [Decreto n. 256].
https://goo.gl/WxtCiy

Ponemon Institute LLC (2015). "Fifth annual benchmark study on privacy & security of healthcare data".
http://media.scmagazine.com/documents/121/healthcare_privacy_security_be_30019.pdf

Punter, Teade; **Solingen, Rini-Van**; **Trienekens, Jos** (1997). "Software product evaluation – current status and future ne-

eds for customers and industry". *Fourth IT Evaluation Conf.* *http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.39.3442&rep=rep1&type=pdf*

South Africa (2002). "No. 68 of 2002: Electronic communications security". *Government gazette*, v. 452, n. 24356. *http://us-cdn.creamermedia.co.za/assets/articles/attachments/00466_a68-02.pdf*

Symantec (2015). "Internet security threat report". *http://www.symantec.com/security_response/publications/threatreport.jsp*

Tabachnick, Barbara G.; **Fidell, Linda S.** (2006). "Using Multivariate Statistics" (5th ed.). Pearson/Allyn & Bacon. ISBN: 0205459382

Trinidad and Tobago (2011). "Data protection act". *Parliament of Trinidad and Tobago*. Act. n. 13 of 2011. TTO-2011-L-88403. *https://www.ilo.org/dyn/natlex/docs/ELECTRONIC/88403/101074/F1410860608/TTO88403.pdf*

Unión Europea (2016). "Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)". *Diario oficial de la Unión Europea* L119 de 4 de mayo de 2016, Luxemburgo. *https://goo.gl/2lwOM4*

United States (1996). "Health insurance portability and accountability act of 1996". *Department of Health and Human Services, U.S. Government Publishing Office of July 31, 1996*. *http://www.hhs.gov/ocr/privacy/hipaa/administrative/statute/hipaastatutepdf.pdf*

Villalba-de-Benito, María-Teresa; **De-Buenaga-Rodríguez, Manuel**; **Gachet-Páez, Diego**; **Aparicio-Galisteo, Fernando** (2015). "Security analysis of an IoT architecture for Healthcare". In: *Healthy IoT: 2nd EAI Intl conf on IoT technologies for health care*. Lecture Notes of ICST. Springer. ISSN: 1867-8211 *http://hdl.handle.net/11268/4750*

Comunicación es una lista de distribución en castellano para debatir y estar al día sobre teoría de la comunicación, comunicación política, comunicación industrial, relaciones públicas, comunicación audiovisual y multimedia, radio y televisión, cinematografía, periodismo, periodismo de datos, divulgación de la ciencia, medios y cibermedios, redes sociales... y todos los aspectos relacionados con la COMUNICACIÓN.

Empezó a funcionar en enero de 2017 y está alojada en el servicio de listas de *RedIRIS*, desde donde es posible consultar sus archivos:
https://listserv.rediris.es/cgi-bin/wa?A0=COMUNICACION

La lista cuenta con 2 moderadores que permanentemente filtran los mensajes para evitar spam, mensajes inapropiados, anuncios, mensajes repetidos, etc.:

Isabel Olea (*EPI*, León)
Tomàs Baiget (*EPI*, Barcelona)

Puedes suscribirte a ***Comunicación*** en:
https://listserv.rediris.es/cgi-bin/wa?SUBED1=COMUNICACION&A=1